

Optimisation de plans d'expérience par méthodes particulières

Billy Amzal^{*,*,***} — Eric Parent^{*} — Frédéric Y. Bois^{**} — Christian P. Robert^{***}**

^{*} ENGREF, Laboratoire GRESE
19 avenue du Maine
75732 Paris Cedex 15
amzal@engref.fr

^{**} INERIS, Unité de Toxicologie Expérimentale
Parc Technologique Alata, BP2, 60550 Verneuil-en-Halatte

^{***} Université Paris Dauphine, CEREMADE

RÉSUMÉ. Nous proposons un nouvel algorithme stochastique de recherche de plan d'expérience optimal bayésien, dans le cadre de modèles non linéaires et de dimension élevée. A la manière de Peter Müller, nous transformons le problème d'optimisation en un problème de simulation de chaînes de Markov qui explorent la surface de l'espérance du critère d'optimisation. Le plan d'expérience optimal correspond alors au mode de cette surface vue comme une distribution de probabilité. Pour être performant dans les cas complexes, notre algorithme combine l'approche « particulière » pour explorer efficacement la surface et un effet de recuit simulé pour concentrer les simulations près des modes. Il est testé ici sur un exemple de changements multiple sur une série temporelle.

ABSTRACT. We propose a new stochastic algorithm for Bayesian optimal design in nonlinear and high dimensional models. Like in the recent work of Peter Müller, we turn the optimization problem into a matter of Monte Carlo Markov chains simulations to explore the expected utility surface. The optimal design is then the mode of this surface seen as a probability distribution. Our algorithm mixes a “particles” method to efficiently explore high dimensional multimodal surfaces, with simulated annealing to concentrate the samples near the modes. We test it in a multiple change-point problem for time series data.

MOTS-CLÉS : optimisation stochastique, planification expérimentale, théorie de la décision bayésienne, algorithmes MCMC, méthodes particulières, recuit simulé.

KEYWORDS: stochastic optimization, experimental designs, Bayesian decision theory, MCMC algorithms, particles methods, simulated annealing.

1. Introduction

L'optimisation des plans d'expérience pour des modèles non linéaires et de grande dimension reste un problème d'actualité à cause de sa complexité et de sa récurrence dans les applications. Avec les récents progrès des moyens de calcul, de nouvelles méthodes d'optimisation stochastique ont vu le jour (Robert, 1996). L'approche de la statistique classique a atteint ses limites pour des modèles non triviaux où aucun calcul explicite n'est possible. Elle ne permet pas de considérer les critères d'optimalité d'un point de vue stochastique, ni d'intégrer simplement les coûts et les contraintes dans l'optimisation. En revanche, l'approche bayésienne se trouve naturellement adaptée à ce type d'optimisation stochastique, puisqu'elle permet de l'inscrire dans un contexte de théorie statistique de la décision. Ainsi, on ramène la recherche d'un optimum sous contraintes à un problème de simulation autour d'un mode. Peter Müller a exploité cette idée dans de récents papiers (Müller, 1998), nous rappelons brièvement sa méthode. Puis, nous présentons une nouvelle technique d'optimisation par simulation, de type particulière qui intègre un effet de recuit simulé. Nous l'illustrons par un exemple où les calculs explicites sont simples, de manière à comparer les performances.

2. L'approche de Peter Müller

2.1. Optimisation par méthodes MCMC

Décrivons d'abord le problème de planification d'expérience qui intéressait Peter Müller et qui restera notre cadre de travail dans la suite. Nous désignons par d le plan d'expérience choisi, ou plus généralement une décision prise avant la collecte des données. Nous supposons alors que ces données sont tirées d'une loi $p(y/d, \theta)$ dépendant d'un paramètre θ , de dimension quelconque, et du choix de d . Dans l'approche bayésienne, nous disposons d'une loi *a priori* sur le paramètre θ , notée $p(\theta)$. Pour un jeu de (d, θ, y) , nous sommes en mesure de calculer un gain ou une utilité dite « jointe » $u(d, \theta, y)$, que l'on supposera en outre positive. Le but de l'optimisation est alors de trouver le d^* qui maximise l'utilité moyenne $U(d)$:

En termes probabilistes, $U(d)$ peut être interprétée comme une loi sur les plans d'expérience (à une constante multiplicative près), et d^* comme le mode de cette loi.

$$d^* = \arg \max U(d) = \arg \max (E_{\theta, y}(u(d, \theta, y)))$$

En pratique, il n'est pas possible de calculer $U(d)$. Par contre nous pouvons facilement exprimer, toujours à une constante près, la loi jointe h sur (d, θ, y) :

$$h(d, \theta, y) \propto u(d, \theta, y)p(\theta, y/d) = u(d, \theta, y)p(y/d, \theta)p(\theta)$$

Peter Müller propose alors d'implémenter un algorithme de Métropolis-Hastings pour simuler h . Le tirage indépendant selon $p(\theta, y/d)$ pris pour loi instrumentale sur (θ, y) permet écrire facilement le taux d'acceptation en fonction uniquement des

utilités jointes et de la loi instrumentale sur les plans d'expérience. Au cours de la simulation, on retient le plan qui maximise u .

2.2. Effet de recuit simulé

Dans les applications, l'espace des plans d'expérience est de grande dimension, et la fonction U a une très faible courbure autour de son mode. De plus, les performances de l'algorithme dépendent beaucoup de la variance de la loi $p(\theta, y/d)$. Il est donc judicieux de rendre la surface U « plus pointue », en simulant des plans d'expériences d qui suivent marginalement non plus U mais U^J où J est un entier qui croît avec le pas de l'algorithme. On retrouve donc les mêmes commodités que précédemment, mais la loi simulée est maintenant :

$$h_j(d, \theta_1, \dots, \theta_J, y_1, \dots, y_n) \propto \prod_{j=1}^J u(d, \theta_j, y_j) p(\theta_j, y_j / d)$$

3. Optimisation par méthodes particulières

A la lumière des récents progrès des méthodes MCMC particulières (Doucet *et al.*, 2001), nous proposons une amélioration de l'algorithme précédent, surtout valable dans les cas les plus complexes de grande dimension, ou pour des surfaces U multimodales par exemple. Nous ne décrivons pas ici le détail de l'algorithme faute de place, mais nous en présentons l'intuition. L'idée est de simuler non plus une mais plusieurs chaînes de Markov simultanément, et qui peuvent interagir les unes avec les autres. On sélectionne ainsi les chaînes en multipliant celles proches des modes, et en éliminant celles qui en sont loin. Cette procédure de sélection peut être de type SIR (« sampling importance resampling ») par exemple. Comme précédemment, on augmente la puissance de la loi simulée. A chaque pas de l'algorithme, on a alors non plus un tirage mais un échantillon de h_j . Lorsque $J \rightarrow +\infty$, l'échantillon de plans d'expériences obtenu se rapproche de celui d'un Dirac en d^* . L'intérêt de la sélection est de pallier l'inconvénient classique de l'algorithme de Métropolis-Hastings à bloquer les simulations autour d'un mode local, et d'explorer plus efficacement les surfaces d'utilité en grande dimension.

4. Application à un problème de points de changement sur une série temporelle

4.1. Énoncé du problème

Nous testons maintenant l'algorithme sur un modèle de série temporelle avec changements multiples. Il s'agit d'un exemple académique où les calculs explicites sont possibles de manière à mesurer les performances de la méthode proposée. Ainsi, nous considérons une série de données gaussiennes sur un intervalle de longueur T , avec n changements indépendants de ses moyenne μ_i et variance σ_i^2 . Nous disposons de lois *a priori* sur $\theta = (\mu_1, \dots, \mu_{n+1}, \sigma_1, \dots, \sigma_{n+1})$. La décision consiste ici en un choix

des n temps de changement $d = (t_1, \dots, t_n)$ entre 0 et T . Le modèle s'écrit $(y(t) | \theta, d) \sim N(\mu_i, \sigma_i^2)$ pour $t_i \leq t < t_{i+1}$. Nous proposons de chercher les temps qui laissent les moyennes estimées le plus stable possible entre deux temps, c'est-à-dire que nous voulons minimiser le coût suivant :

$$C(d) = E_{\theta, y} \left[\sum_{i=0}^n (\hat{\mu}_i - \hat{\mu}_{i+1})^2 \right] \text{ où } \hat{\mu}_i = \frac{1}{t_{i+1} - t_i} \sum_{j=t_i}^{t_{i+1}-1} y(j) \text{ avec } t_0 = 0 \text{ et } t_{n+1} = T$$

4.2. Résultats comparés

Ce problème est soluble explicitement, par un conditionnement de l'espérance et une méthode d'optimisation par multiplicateurs de Lagrange. Nous comparons alors les performances de notre algorithme à 110 particules à celles de 110 algorithmes de Müller indépendants simultanés, pour 10 temps de rupture parmi 180. La figure 1 montre que l'interaction des chaînes permet de simuler en moyenne des plans bien moins coûteux. Bien sûr, cela est accentué par d'éventuels doublons de particules provenant du rééchantillonnage. Mais la figure 2 montre que notre algorithme permet des simulations bien plus proches de l'optimum.

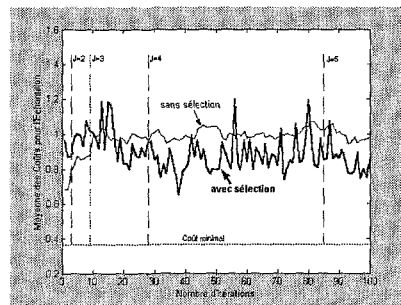


Figure 1. Moyenne des coûts sur les 110 particules à chaque itération

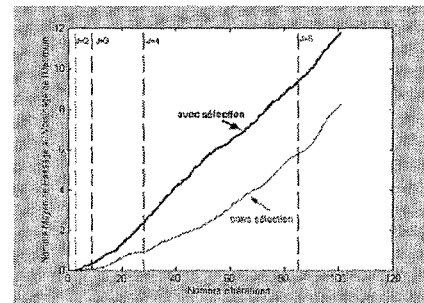


Figure 2. Nombre moyen de passages du meilleur coût sur les 110 dans un voisinage (de 20 %) du coût minimal

5. Bibliographie

- Doucet A., Gordon N., de Freitas N., (Eds) *Sequential Monte Carlo methods in practice*, New-York, Springer-Verlag, 2001
- Muller P., « Simulation based optimal design », *Bayesian Statistics 6*, Oxford University Press, 1999, pp 459-474
- Robert C.P., *Methodes de Monte Carlo par chaines de Markov*, Paris, Economica, 1996